

## **ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI INDEK MUTU BENANG MENGGUNAKAN METODE POHON REGRESI (Studi Kasus di PT. Industri Sandang Nusantara Unit Patal Grati)**

**Hesti Sari Dewi K. W.<sup>1</sup>, Yuciana Wilandari<sup>2</sup>, Sudarno<sup>2</sup>**

<sup>1</sup>Alumni Program Studi Statistika FMIPA Universitas Diponegoro

<sup>2</sup>Staf Pengajar Program Studi Statistika FMIPA Universitas Diponegoro

### **Abstract**

Quality for ripe material (yarn) really necessary for the company, therefore needs to control the product (ripe material), so we are able to know the unmatched product percentage of the company standard and to know the cause of the unmatched. The appropriate method to know the influential factor to company yarn quality successes, among those regression tree method. Regression tree is one of CART's classification method. CART is a useful non parametric statistical method to get an accurate data group as distinguishing as of a classification. Because it has continuous type of response variable, so that CART can create regression tree. Regression tree is utilized to figure relationship among one response variable with one or more predictor variable that gets continued character and also category. The variables that have influence for yarn quality index at PT. Industri Sandang Nusantara Patal Grati's Unit are raw material, machine output year, air humidity (RH) and hall temperature. The result of the research is that the year of machine output variable is the most influence to foot up yarn quality index and has main contribution in the formation of regression tree.

**Keywords :** Regression tree, CART, Yarn quality index, Rayon.

### **1. Pendahuluan**

Perkembangan industri dewasa ini mendorong suatu perusahaan untuk selalu bersaing dengan perusahaan lainnya. Berbagai strategi harus ditempuh untuk memenangkan persaingan pada era pasar bebas. Salah satu strategi yang diambil oleh perusahaan untuk dapat memenangkan persaingan tersebut yaitu dengan membenahi hasil dari produknya agar dapat diterima di pasaran. Untuk membenahi hasil produk tersebut salah satu cara yang umum digunakan perusahaan yaitu mengadakan suatu sistem pengendalian mutu, baik mulai dari pemilihan bahan baku, proses produksi hingga inspeksi hasil akhir dari suatu proses (produk jadi). Karena mutu merupakan salah satu faktor dari proses produksi yang berhubungan langsung dengan kepuasan konsumen. Konsumen akan tertarik pada suatu produk bila produk tersebut sesuai keinginannya, oleh sebab itu diharapkan perusahaan mampu untuk memproduksi suatu produk yang benar-benar sesuai dengan kebutuhan masyarakat luas.

PT. Industri Sandang Nusantara Unit Patal Grati merupakan salah satu jenis perusahaan *job order* dimana produksi dilakukan tergantung pada pesanan konsumen. Dengan adanya produk yang kurang sesuai dengan keinginan konsumen dan adanya permintaan dari luar negeri maka bagian pengendalian mutu pada perusahaan ini mengadakan pemeriksaan yang ketat mulai dari awal pemilihan bahan baku, pengendalian proses hingga pemeriksaan pada produk jadi. Mutu terhadap bahan jadi (benang) sangat penting bagi perusahaan, sehingga perlu diadakan pengawasan terhadap produk tersebut agar dapat diketahui prosentase produk yang tidak sesuai dengan standar perusahaan dan mengetahui penyebab produk tersebut tidak sesuai dengan standar yang ditentukan.

Metode yang digunakan untuk mengetahui faktor-faktor yang berpengaruh terhadap keberhasilan mutu benang perusahaan, diantaranya adalah metode pohon regresi. Metode ini merupakan salah satu dari metode klasifikasi pohon (*Classification and Regression Trees / CART*). Metode pohon regresi memiliki banyak kelebihan dibandingkan metode lain, beberapa kelebihanannya antara lain dapat melakukan eksplorasi data yang kompleks, mengidentifikasi berbagai variabel prediktor yang mempunyai hubungan struktural dengan variabel responnya dan memprediksi nilai respon<sup>[3]</sup>. Kelebihan lain dari pohon regresi adalah kemampuannya dalam mendeteksi interaksi antar variabel secara lokal atau bekerja untuk menemukan subgroup data yang bermakna serta interpretasi hasilnya lebih mudah daripada proses regresi biasa, karena identifikasi adanya pengaruh dari variabel prediktor dalam pohon regresi dilakukan dalam masing-masing subgroup data bukan dalam keseluruhan data seperti halnya regresi biasa. Dalam penelitian ini variabel-variabel yang diduga berpengaruh terhadap keberhasilan mutu produksi benang (indek mutu benang) khususnya jenis 30/1 rayon di PT. Industri Sandang Nusantara Unit Patal Grati adalah bahan baku, tahun keluaran mesin, kelembaban udara (RH), dan temperatur ruangan.

## 2. Bahan dan Metode

### 2.1. Metode Pohon Regresi

Pengelompokkan objek ke dalam satu atau beberapa kelompok berdasarkan variabel yang diamati disebut klasifikasi. Salah satu metode yang digunakan dalam masalah pengklasifikasian adalah metode klasifikasi pohon atau *Classification And Regression Trees* (CART). Metode ini digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Metode ini dikenal juga sebagai metode pemilahan rekursif secara biner (*binary recursive partitioning*)<sup>[1][3]</sup>, artinya sekelompok data yang terkumpul dalam suatu ruang yang disebut simpul (*node*) dapat dipilah menjadi dua simpul anak dan setiap simpul anak dapat dipilah lagi menjadi dua simpul anak. Demikian seterusnya, hingga diperoleh suatu simpul akhir (*terminal node*) dan berhenti jika telah memenuhi kriteria tertentu. Pohon yang dihasilkan bergantung pada skala variabel respon. Jika data variabel respon kontinu maka pohon yang dihasilkan adalah pohon regresi. Variabel yang berpengaruh dalam pohon regresi adalah variabel yang menentukan adanya pemilahan (*splitting*) bagi ruang variabel prediktor.

Pembentukan pohon regresi memerlukan 4 komponen yaitu:

1. Satu gugus pertanyaan dikotomis dalam bentuk “Apakah  $x \in A$ ” dengan  $x$  merupakan suatu amatan sampel dan  $A \subset X$  (ruang variabel prediktor), dimana jawaban dari pertanyaan tersebut menentukan pemilahan bagi ruang variabel prediktor. Amatan dengan jawaban ‘ya’ masuk ke anak ruang  $A$ , sedangkan amatan dengan jawaban ‘tidak’ masuk ke ruang komplemen  $A$ . Anak ruang sampel yang terbentuk disebut simpul.
2. Kriteria *goodness of split*  $\Delta R(s^*, t)$  merupakan alat evaluasi bagi pemilahan yang dilakukan oleh pemilah  $s$  pada simpul  $t$ .
3. Kriteria yang digunakan untuk menentukan ukuran pohon yang optimal atau tepat yaitu berdasarkan sampel uji atau sampel validasi silang.
4. Statistik digunakan sebagai ringkasan atau kesimpulan dari tiap simpul akhir. Pada analisis pohon regresi digunakan statistik rata-rata respon sebagai nilai dugaan pada setiap simpul terminal.

Proses pembentukan pohon regresi dimulai dari simpul utama atau simpul akar. Pada simpul akar terdapat suatu kelompok data, pemilahan pada simpul akar berdasarkan variabel terpenting. Selanjutnya simpul utama dipilah menjadi simpul anak kiri dan simpul anak kanan. Kemudian kedua simpul anak tersebut dipilah kembali menjadi dua.

### 2.1.1. Pertumbuhan (*Growing*) Pohon Regresi

Pohon regresi terbentuk dari hasil pemilahan data pada setiap simpul induk menjadi dua simpul anak. Aturan pemilahannya adalah sebagai berikut<sup>[1]</sup>:

1. Tiap pemilahan bergantung pada satu nilai pemilah yang hanya berasal dari satu variabel prediktor.
2. Untuk variabel prediktor kontinu  $X_j$ , pemilahan berasal dari pertanyaan “Apakah  $x_j \leq c_i$ ?” untuk  $c_i \in \mathcal{R}^j$  ( $\mathcal{R}^j$  merupakan ruang sampel dari variabel  $X_j$ ). Sehingga, jika ruang sampelnya berukuran  $N$  dan terdapat sebanyak-banyaknya  $n$  nilai amatan berbeda pada variabel  $X_j$ , maka akan terdapat sebanyak  $n-1$  pemilahan yang berbeda yang dibentuk oleh gugus pertanyaan dengan  $i = 1, 2, \dots, n-1$  dan  $c_i$  adalah nilai tengah antara dua nilai amatan variabel  $X_j$  berukuran berbeda.
3. Untuk variabel prediktor yang bersifat kategori, pemilahan yang terjadi berasal dari semua kemungkinan pemilahan berdasarkan terbentuknya dua anak gugus yang saling lepas. Apabila  $X_j$  merupakan variabel kategori berskala nominal bertaraf  $L$ , maka akan terdapat  $2^{L-1}-1$  kemungkinan pemilahan. Sedangkan jika  $X_j$  merupakan variabel kategori berskala ordinal maka akan terdapat  $L - 1$  pemilahan yang mungkin.

Pohon regresi dibentuk melalui suatu pemilahan rekursif berdasarkan aturan pemilahan tersebut. Proses pemilahan dilakukan pada tiap simpul untuk membentuk pohon regresi dengan aturan sebagai berikut:

1. Mencari semua kemungkinan pemilahan pada tiap variabel prediktor berdasarkan aturan pemilahan tersebut.
2. Memilih pemilahan yang terbaik pada masing-masing variabel prediktor, kemudian melakukan pemilihan pemilahan yang terbaik dari kumpulan pemilahan terbaik tersebut.

Pemilahan terbaik (*goodness of split*) adalah pemilahan yang memaksimalkan ukuran kehomogenan di dalam masing-masing simpul anak relatif terhadap simpul induknya dan yang memaksimalkan ukuran pemisahan antara kedua simpul anak tersebut. Pemilahan terbaik dihitung berdasarkan selisih jumlah kuadrat deviasi antara simpul induk dengan kedua simpul anak pemilahnya. Selisih terbesar akan dijadikan sebagai pemilahan terbaik. Misalkan diketahui simpul  $t$  berisi sampel  $\{(x_{n(t)}, y_{n(t)})\}$ , sedangkan  $n(t)$  adalah jumlah amatan dalam simpul  $t$  maka rata-rata respon dalam simpul  $t$  adalah<sup>[1]</sup>:

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_{n(t)} \in t} y_{n(t)} \quad (1)$$

dengan  $y_{n(t)}$  = nilai individu atau amatan variabel respon dalam simpul  $t$

$n(t)$  = jumlah amatan dalam simpul  $t$

sehingga jumlah kuadrat deviasi yang digunakan sebagai kriteria kehomogenan pada suatu simpul  $t$  adalah

$$R(t) = \sum_{x_{n(t)} \in t} (y_{n(t)} - \bar{y}(t))^2 \quad (2)$$

Jika terdapat pemilah  $s$  yang memilah  $t$  menjadi simpul anak kiri  $t_L$  dan simpul anak kanan  $t_R$ , maka kriteria selisih jumlah kuadrat deviasi adalah

$$\Delta R(s, t) = R(t) - (R(t_L) + R(t_R)) \quad (3)$$

dengan  $R(t)$  = Jumlah kuadrat deviasi suatu simpul  $t$

$R(t_L)$  = Jumlah kuadrat deviasi suatu simpul anak kiri  $t_L$

$R(t_R)$  = Jumlah kuadrat deviasi suatu simpul anak kanan  $t_R$

dan pemilah terbaik  $s^*$  dari  $t$  adalah pemilahan pada  $S$  yang sedemikian hingga

$$\Delta R(s^*, t) = \max_{s \in S} \Delta R(s, t) \quad (4)$$

dengan  $S$  merupakan gugus yang berisi semua kemungkinan pemilahan

Pohon regresi dibentuk melalui pemilahan simpul secara rekursif, yaitu dengan memaksimumkan fungsi  $\Delta R(s, t)$  tersebut.

### 2.1.2. Penghentian Pembentukan Pohon Regresi

Proses pembentukan pohon regresi akan berhenti apabila sudah tidak dimungkinkan lagi dilakukannya proses pemilahan. Proses pemilahan akan berhenti apabila hanya terdapat satu amatan yang ada di dalam simpul terakhir atau adanya batasan minimum  $n$ . Pengembangan pohon berhenti apabila pada simpul terdapat  $n_i < 5$ . Semua amatan yang berada di dalam sebuah simpul merupakan anggota nilai respon yang relatif homogen. Pohon regresi yang terbentuk sebagai hasil dari proses ini dinamakan pohon regresi maksimal atau *Largest tree* ( $T_{\max}$ )<sup>[3]</sup>.

### 2.1.3. Pemangkasan (*Pruning*) Pohon Regresi

Pohon regresi yang dibentuk melalui proses pemilahan secara rekursif akan berukuran sangat besar. Hal ini disebabkan karena aturan penghentian yang digunakan hanya berdasarkan banyaknya amatan pada simpul terminal atau besarnya penurunan tingkat keragaman dalam tiap simpul anak hasil pemilahan. Semakin banyak pemilahan yang dilakukan maka tingkat kesalahan prediksi juga akan semakin kecil. Namun pohon regresi yang terbesar atau maksimal terlalu sulit untuk dipahami sehingga menyebabkan *overfitting* untuk data baru. Masalah tersebut diatasi dengan melakukan pemangkasan pada pohon regresi maksimal untuk mendapatkan pohon regresi dengan ukuran yang optimal<sup>[3]</sup>.

Langkah awal pemangkasan dilakukan terhadap  $T_1$ , yaitu suatu subpohon dari pohon terbesar  $T_{\max}$ . Untuk mendapatkan  $T_1$  dari  $T_{\max}$ , diambil  $t_L$  dan  $t_R$  yang merupakan simpul anak kiri dan simpul anak kanan dari  $T_{\max}$  yang dihasilkan dari pemilahan pada setiap simpul induk  $t$ . Karena  $R(t) \geq R(t_L) + R(t_R)$ , maka ketika terdapat dua simpul anak dan simpul induk yang memenuhi persamaan  $R(t) = R(t_L) + R(t_R)$ , simpul anak  $t_L$  dan  $t_R$  tersebut dipangkas. Proses ini diulangi sampai tidak memungkinkan lagi dilakukan pemangkasan. Proses pemangkasan kompleksitas kesalahan terkecil adalah pemotongan hubungan terlemah pada pohon regresi. Untuk sembarang  $T_t$  yang merupakan cabang dari  $T_1$ , besar rata-rata kuadrat deviasi didefinisikan dengan

$$R(T_t) = \sum_{t' \in T_t} R(t') \quad (5)$$

dengan  $R(t') =$  rata-rata kuadrat deviasi pada simpul  $t$

$$R(t') = \frac{1}{n(t)} \sum_{x_{n(t)} \in t} (y_{n(t)} - \bar{y}(t))^2 \quad (6)$$

untuk setiap simpul  $t \in T_1$ , didefinisikan  $\{t\}$  sebagai suatu sub cabang dari  $T_t$  yang hanya terdiri dari satu simpul. Ukuran kompleksitas kesalahan dari sub cabang  $\{t\}$  adalah

$$R_\alpha(\{t\}) = R(t') + \alpha \quad (7)$$

dan ukuran kompleksitas kesalahan dari cabang atau pohon  $T_t$  adalah

$$R_\alpha(T_t) = R(T_t) + \alpha |\tilde{T}_t| \quad (8)$$

dengan  $\alpha =$  Parameter kompleksitas mengenai kesalahan bagi penambahan satu simpul akhir pada pohon  $T_t$

$$|\tilde{T}_t| = \text{Banyaknya simpul akhir yang dimiliki pohon } T_t$$

Nilai  $\alpha$  merupakan suatu parameter kompleksitas mengenai kesalahan bagi penambahan satu simpul akhir pada pohon  $T_t$ . Semakin besar nilai  $\alpha$  maka ukuran pohon yang dihasilkan akan kecil, sebaliknya jika nilai  $\alpha$  kecil maka kompleksitas simpulnya juga kecil sehingga ukuran pohon yang dihasilkan akan besar. Sehingga contoh sebuah  $T_{\max}$  memiliki simpul akhir yang berisi hanya satu objek atau homogen sehingga nilai  $\alpha$  untuk  $T_{\max}$  adalah 0 dan nilai ukuran kompleksitas mencapai minimum karena  $R(T_{\max}) = 0$ . Nilai  $\alpha$  akan terus meningkat selama proses pemangkasan berlangsung dan akan mencapai nilai terbesar pada saat simpul akhir = simpul akar.

Nilai kompleksitas pemangkasan menentukan pohon bagian  $(T_t)(\alpha)$  yang meminimumkan  $R_\alpha(T_t)$  pada seluruh pohon bagian untuk setiap nilai  $\alpha$ . Nilai parameter kompleksitas  $\alpha$  akan secara perlahan meningkat selama proses pemangkasan. Selanjutnya pencarian pohon bagian  $T(\alpha) < T_{\max}$  yang dapat meminimumkan  $R_\alpha(T_t)$  yaitu:

$$R_\alpha((T_t(\alpha))) = \min_{T < T_{\max}} R_\alpha(T_t) \quad (9)$$

Hasil dari proses pemangkasan adalah deretan sub pohon dengan ukuran yang semakin mengecil, yaitu  $T_1 > T_2 > \dots > \{t_1\}$ , artinya pohon  $T_1$  adalah induk bagi pohon  $T_2$ , pohon  $T_2$  adalah induk bagi pohon  $T_3$ , demikian seterusnya dengan deretan  $\alpha$  dalam urutan meningkat, yaitu  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$  dengan  $\alpha_k < \alpha_{k+1}$  untuk  $k \geq 1$ , dan  $\alpha_k = 0$  untuk  $k = 1$ .

#### 2.1.4. Penentuan Ukuran Pohon Regresi Yang Optimal

Ukuran pohon regresi yang besar akan menyebabkan nilai kompleksitas kesalahan yang tinggi, tetapi semakin besar pohon regresi maka tingkat kesalahan prediksinya juga akan semakin kecil, sehingga perlu dipilih pohon regresi optimal yang berukuran sederhana tetapi juga memberikan nilai kesalahan prediksi yang cukup kecil<sup>[5]</sup>. Pohon regresi optimal dinotasikan dengan  $T_k$  yaitu sebuah sub pohon terkecil dari  $T_{\max}$  yang meminimumkan tingkat kesalahan prediksi. Ada beberapa cara yang digunakan untuk menduga tingkat kesalahan prediksi dari suatu model pohon regresi<sup>[1]</sup>, salah satunya adalah menggunakan penduga validasi silang lipat V (*Cross Validation Estimate*). Keseluruhan data akan digunakan untuk membentuk deretan pohon  $\{T_k\}$  dan deretan parameter kompleksitas  $\{\alpha_k\}$ . Untuk memperoleh nilai penduga validasi silang lipat V, amatan induk  $\mathcal{L}$  yang berukuran N dibagi secara acak menjadi V kelompok, yaitu  $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_V$ , yang berukuran sama yaitu n. Sampel *learning* ke-v adalah  $\mathcal{L}^{(v)} = \mathcal{L} - \mathcal{L}_v$ ,  $v = 1, 2, \dots, V$  yang digunakan untuk memperoleh nilai penduga  $d_k^{(v)}(x)$ . Penduga validasi silang lipat V dirumuskan dengan

$$R^{CV}(T_k) = \frac{1}{N} \sum_{v=1}^V \sum_{(x_n, y_n) \in \mathcal{L}_v} [y_n - d_k^{(v)}(x_n)]^2 \quad (10)$$

dengan  $y_n$  = Nilai amatan variabel respon dari data learning ke-V

$d_k^{(v)}(x_n)$  = Nilai penduga variabel respon dari data learning ke-V sebagai ukuran tingkat kesalahan prediksi

Pohon regresi optimal adalah  $T_{k0}$  yang memenuhi kriteria

$$R^{CV}(T_{k0}) = \min_k R^{CV}(T_k) \quad (11)$$

#### 2.1.5. Ketelitian Dugaan Pohon Regresi

Ketelitian dugaan pohon regresi digunakan untuk mengetahui seberapa besar pengaruh hasil prediksi terhadap amatannya atau kecocokan pohon regresi. Ukuran yang

digunakan untuk melihat ketelitian dugaan terbaik suatu pohon regresi diantaranya dengan melihat kriteria *MSE* terkecil<sup>[2]</sup>.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (12)$$

## 2.2. Variabel Penelitian

Variabel yang digunakan pada penelitian ini meliputi variabel respon yaitu indeks mutu benang (IMB) di Patal Grati dengan skala 300-500 dari beberapa jenis nomer benang dan 4 variabel prediktor yaitu bahan baku ( $X_1$ ), tahun keluaran mesin ( $X_2$ ), kelembaban udara/RH ( $X_3$ ) dan temperatur ruangan ( $X_4$ )

Variabel-variabel yang digunakan dalam penelitian terangkum dalam Tabel 1

**Tabel 1.** Variabel Penelitian

Variabel Penelitian	Kategori
Y = Nilai IMB	Kontinu (skala 300-500)
$X_1$ = Bahan baku (BHNBU)	Kontinu (skala 59-91)
$X_2$ = Tahun Keluaran Mesin (THKELMSN)	1: Tahun 1961 (Plat) 2: Tahun 1979 (Hawa Lama) 3: Tahun 1995 (Zinser) 4: Tahun 1997 (Hawa Baru)
$X_3$ = Kelembaban udara (RH)	1: Kering ( $< 58^0\text{C}$ ) 2: Standar ( $60 \pm 2^0\text{C}$ ) 3: Lembab ( $> 62^0\text{C}$ )
$X_4$ = Temperatur ruangan (TEMPERTR)	1: Rendah ( $< 28^0\text{C}$ ) 2: Standar ( $30 \pm 2^0\text{C}$ ) 3: Tinggi ( $> 32^0\text{C}$ )

## 2.3. Metode Analisis

Langkah-langkah analisis yang dilakukan dalam mencari bentuk pohon regresi terhadap indeks mutu benang di Patal Grati dengan metode CART (*Classification and Regression Tree*) dengan menggunakan *software* CART versi 6.0 dan dilakukan dengan adalah sebagai berikut:

- i. Data dibagi menjadi dua yaitu data *learning* dan data *testing*. Data *learning* digunakan untuk proses pembentukan pohon regresi, sedangkan data *testing* digunakan untuk melihat kemampuan pohon regresi yang terbentuk dalam menduga atau memprediksi data baru.
- ii. Pembentukan pohon regresi yaitu dengan tahapan berikut:
  - a. Pertumbuhan pohon regresi maksimal, diawali dengan pemilahan yang dilakukan dengan cara memilah data hasil pengujian indeks mutu benang sesuai aturan pertumbuhan dan aturan pemilahan. Pemilahan terbaik diantara variabel prediktor ditentukan berdasarkan selisih terbesar jumlah kuadrat deviasi antara simpul induk dan kedua simpul anak yang memilahnya.
  - b. Penghentian pemilahan pohon regresi akan berhenti apabila sudah tidak dimungkinkan lagi dilakukan pemilahan. Hal ini dikarenakan hanya terdapat satu amatan yang ada dalam simpul akhir atau adanya batasan minimum  $n$ .
  - c. Pemangkasan pohon regresi adalah proses pemangkasan pohon yang dimulai dengan memangkas pohon regresi maksimal sampai diperoleh ukuran pohon regresi yang paling kecil. Pemangkasan dengan menggunakan kriteria selisih

- jumlah kuadrat deviasi antara simpul induk dengan jumlah kedua simpul anak. Dari hasil proses pemangkasan ini diperoleh deretan pohon regresi dengan ukuran yang semakin mengecil, dan tingkat kesalahan prediksi yang semakin membesar.
- d. Pemilihan pohon regresi optimal dilakukan dengan kriteria kesalahan relatif penduga validasi silang lipat 10 yang minimum.
  - e. Pengukuran ketelitian dugaan dengan memasukkan data *testing* untuk melakukan validasi pohon.

### 3. Hasil dan Pembahasan

Jumlah data keseluruhan yang digunakan dalam penelitian ini adalah 222. Data penelitian dibagi menjadi dua yaitu data *learning* dan data *testing*. Data *learning* diambil sebanyak 211 (95%) digunakan untuk proses pertumbuhan pohon regresi, sedangkan data *testing* diambil sebanyak 11 (5%) digunakan untuk melakukan validasi model, yaitu melihat kemampuan dari model yang terbentuk dalam menduga atau memprediksi jumlah indek mutu benang dari data baru<sup>[1]</sup>.

#### 3.1. Pertumbuhan Pohon Regresi Maksimal

Pembentukan pohon regresi dilakukan menurut aturan pemilahan dan aturan pertumbuhan, yaitu dimulai dari pemilahan data indek mutu benang dari 211 sampel oleh variabel pemilah terbaik (diantara variabel bahan baku, tahun keluaran mesin, kelembaban udara (RH) dan temperatur udara). Tiap-tiap variabel prediktor memiliki jumlah kemungkinan pemilahan yang berbeda, yaitu:

1. Variabel bahan baku memiliki 15 kemungkinan pemilahan.
2. Variabel tahun keluaran mesin memiliki 3 kemungkinan pemilahan.
3. Variabel kelembaban udara (RH) memiliki 2 kemungkinan pemilahan.
4. Variabel temperatur ruangan memiliki 2 kemungkinan pemilahan.

Pertumbuhan pohon regresi maksimal dimulai pada data indek mutu benang dari 211 sampel yang terkumpul dalam suatu himpunan yang disebut simpul akar dan diberi nama simpul 1. Masing-masing kemungkinan pemilahan ini akan menghasilkan 2 kelompok data yaitu simpul anak kiri dan simpul anak kanan, kedua simpul anak tersebut diberi nama simpul 2 dan simpul 3. Pemilahan ini diharapkan dapat memaksimumkan ukuran kehomogenan di dalam masing-masing simpul anak relatif terhadap simpul induknya, dan memaksimumkan ukuran pemisahan antara simpul anak kiri dan simpul anak kanan yang terbentuk tersebut.

Sebelum dilakukan proses pemilahan simpul, terlebih dahulu memilih variabel pemilah terbaik dari ke empat variabel prediktor. Pemilihan variabel pemilah terbaik berdasarkan kriteria *goodness of split*. Suatu *split s* akan digunakan untuk memilah simpul  $t$  menjadi dua buah simpul yaitu simpul  $t_R$  dan simpul  $t_L$  jika  $s$  memaksimumkan nilai  $\Delta R(s^*, t) = \max_{s \in S} \Delta R(s, t)$ . Pemilah terbaik yang terpilih untuk simpul 1 (simpul akar) adalah

pemilahan oleh variabel tahun keluaran mesin = 1 (Plat). Variabel tahun keluaran mesin terpilih karena memiliki nilai *improvement* tertinggi dari *competitor* lain. Hal ini dapat dilihat pada Tabel 2.

**Tabel 2.** Kriteria Pemilahan Terbaik Simpul Akar

Main	Competitor	Split	Improvement	N Left	N Right	N Missing
1	THKELMSN	1.0	458.02514	45	166	0
2	BHNBKU	88.5	73.09824	160	51	0
3	RH	1.0	12.29469	97	114	0
4	TEMPERTR	3.0	5.29739	57	154	0

Setelah terbentuk dan terpilih kriteria pemilah terbaik, maka simpul akar yang berisi 211 objek data dipilah menjadi dua buah simpul. Pada masing-masing simpul anak yang terbentuk, jumlah sampel dan banyaknya kategori variabel akan berkurang, sehingga jumlah kemungkinan pemilahan juga akan semakin berkurang. Langkah selanjutnya adalah mengulangi proses pemilahan pada kedua simpul anak yang terbentuk, demikian seterusnya sampai tidak memungkinkan lagi dilakukan pemilahan. Suatu simpul tidak dipilah lagi jika pada simpul tersebut tidak terpenuhi aturan pemilahan dan aturan pertumbuhan. Kriteria penghentian pembentukan pohon regresi ditentukan berdasarkan jumlah amatan dalam simpul terakhir yang hanya tersisa satu amatan saja yang ada dalam simpul terakhir atau semua amatan yang berada dalam sebuah simpul merupakan anggota dari rataan respon yang relatif homogen.

### 3.2. Proses Penghentian Pemilahan Pohon Regresi

Proses pemilahan simpul yang berulang-ulang akan berhenti dan menghasilkan pohon maksimal. Proses pemilahan simpul berhenti karena pada ujung pohon regresi terdapat simpul akhir yang hanya memiliki satu amatan. Pohon regresi maksimal yang diperoleh pada penelitian ini memiliki 34 simpul dalam dan 35 simpul akhir, dengan kedalaman pohon regresi sebesar 10 tingkatan. Kedalaman pohon menunjukkan jumlah level atau tingkatan dari pohon regresi yang dihitung dari simpul akar sampai pada simpul akhir terbawah.

Pada pohon regresi maksimal yang terbentuk, semua variabel prediktor yang diduga mempengaruhi indek mutu benang masuk ke dalam pohon. Variabel yang paling berpengaruh terhadap besarnya indek mutu benang adalah variabel tahun keluaran mesin. Tabel 3 menunjukkan bahwa variabel tahun keluaran mesin paling sering digunakan sebagai pemilah utama maupun pemilah pengganti pada pohon regresi maksimal. Jadi, variabel tahun keluaran mesin merupakan variabel yang mempunyai kontribusi utama dalam pembentukan pohon regresi maksimal disusul oleh variabel bahan baku, temperatur ruangan dan kelembaban udara.

**Tabel 3.** Skor Variabel Penting pada Pohon Regresi Maksimal

Variabel	Skor
Tahun keluaran mesin	100.00
Bahan Baku	15.75
Temperatur Ruangan	2.30
Kelembaban udara	1.80

### 3.3. Proses Pemangkasan Pohon Regresi

Pohon regresi yang dibentuk melalui proses pemilahan secara rekursif berukuran sangat besar. Ukuran pohon regresi yang besar akan menyebabkan kesulitan dalam pemahaman informasi dalam setiap simpul akhir, tetapi semakin besar pohon regresi maka tingkat kesalahan prediksi akan semakin kecil, sehingga perlu dipilih pohon regresi optimal



yang berukuran sederhana tetapi juga memberikan tingkat kesalahan prediksi yang cukup kecil.

Proses pemangkasan pohon regresi maksimal menghasilkan 27 sub pohon. Pemangkasan dimulai dengan mengambil simpul anak kanan  $t_R$  dan simpul anak kiri  $t_L$  dari  $T_{\max}$  yang dihasilkan dari simpul induk  $t$ . Kemudian dilakukan pemotongan cabang yang memiliki selisih jumlah kuadrat deviasi sama dengan 0 atau  $R(t) = R(t_L) + R(t_R)$ . Pemangkasan dilakukan secara iteratif terhadap pohon regresi maksimal sampai kembali diperoleh simpul akar. Gambar 1 adalah salah satu simpul yang akan dipangkas, sedangkan perhitungan pada simpul (*node*) 34 yang akan dipangkas adalah sebagai berikut: Dengan menggunakan Persamaan (1) dan (2), maka pada simpul 34 diperoleh

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_n \in t} y_n = \frac{1}{14} (490 + 480 + 480 + \dots + 480 + 480) = 479.286$$

$$R(t) = \sum_{x_n \in t} (y_n - \bar{y}(t))^2 = \left\{ \begin{aligned} &(490 - 479.286)^2 + (480 - 479.286)^2 + (480 - 479.286)^2 + \dots + \\ &(480 - 479.286)^2 + (480 - 479.286)^2 \end{aligned} \right\}$$

$$= 1892.3$$

Pada simpul akhir (*terminal node*) 33 sebagai simpul anak sebelah kiri

$$\bar{y}(t_L) = \frac{1}{n(t_L)} \sum_{x_n \in t_L} y_n = \frac{1}{13} (490 + 480 + 480 + \dots + 480 + 480) = 479.231$$

$$R(t_L) = \sum_{x_n \in t_L} (y_n - \bar{y}(t_L))^2 = \left\{ \begin{aligned} &(490 - 479.231)^2 + (480 - 479.231)^2 + (480 - 479.231)^2 + \dots + \\ &(480 - 479.231)^2 + (480 - 479.231)^2 \end{aligned} \right\}$$

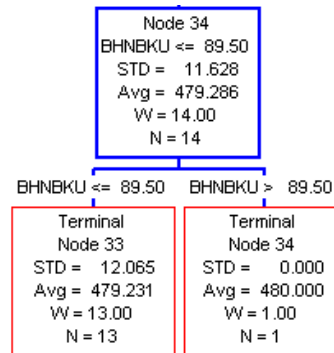
$$= 1892.3$$

Pada simpul akhir (*terminal node*) 34 sebagai simpul anak kanan

$$\bar{y}(t_R) = \frac{1}{n(t_R)} \sum_{x_n \in t_R} y_n = \frac{1}{1} (480) = 480$$

$$R(t_R) = \sum_{x_n \in t_R} (y_n - \bar{y}(t_R))^2 = \{(480 - 480)^2\} = 0$$

Dengan demikian persamaan  $R(t) = R(t_L) + R(t_R)$  terpenuhi untuk simpul 34, sehingga dilakukan pemangkasan.



**Gambar 1.** Simpul 34 pada yang akan dipangkas

### 3.4. Pohon Regresi Optimal

Setelah dilakukan pemangkasan pohon, langkah selanjutnya pemilihan pohon regresi optimal. Dari 27 sub pohon akan dipilih satu pohon regresi dengan nilai penduga pengganti yang terkecil. Dalam mencari pohon regresi digunakan metode penduga validasi silang lipat V (*Cross Validation Estimate*), karena ukuran data yang tidak terlalu besar. Nilai penduga pengganti validasi silang lipat 10 minimum terlihat pada Tabel 4.

**Tabel 4.** Deretan Pohon Regresi

Jumlah Pohon	Simpul Akhir	Kesalahan Relatif Validasi Silang	Kesalahan Relatif Penggantian	Kompleksitas
1	35	$0.43416 \pm 0.08125$	0.27112	0.00000
2	34	$0.43415 \pm 0.08125$	0.27112	0.55008
3	33	$0.43425 \pm 0.08131$	0.27113	1.66664
4	32	$0.43413 \pm 0.08132$	0.27114	2.79506
5	31	$0.43409 \pm 0.08131$	0.27116	3.50113
6	30	$0.43439 \pm 0.08130$	0.27121	10.47010
7	29	$0.43390 \pm 0.08130$	0.27127	12.62624
8	28	$0.43418 \pm 0.08137$	0.27137	19.20636
9	26	$0.43404 \pm 0.08135$	0.27177	40.33329
10	25	$0.43357 \pm 0.08176$	0.27197	41.16729
11	24	$0.42710 \pm 0.08110$	0.27257	120.61566
12	22	$0.42705 \pm 0.08110$	0.27385	129.07173
13	21	$0.42934 \pm 0.08140$	0.27451	133.83493
14	19	$0.42881 \pm 0.08134$	0.27592	142.09959
15	18	$0.42893 \pm 0.08136$	0.27681	178.76660
16	17	$0.42963 \pm 0.08133$	0.27796	232.53529
17	16	$0.43030 \pm 0.08132$	0.27922	254.89468
18	14	$0.43515 \pm 0.08279$	0.28190	270.47086
19	13	$0.43315 \pm 0.08498$	0.28401	424.86075
20	12	$0.43508 \pm 0.08514$	0.28641	485.49503
21	8	$0.42575 \pm 0.08598$	0.29629	498.52921
22	7	$0.41524 \pm 0.08596$	0.30020	789.81628
23	6	$0.40655 \pm 0.09136$	0.30586	1140.83191
24	4	$0.40581 \pm 0.09167$	0.32089	1516.72644
25**	3	$0.37617 \pm 0.08195$	0.36902	9713.99609
26	2	$0.57213 \pm 0.06537$	0.52116	30706.93359
27	1	$0.99275 \pm 0.00741$	1.00000	96643.45313

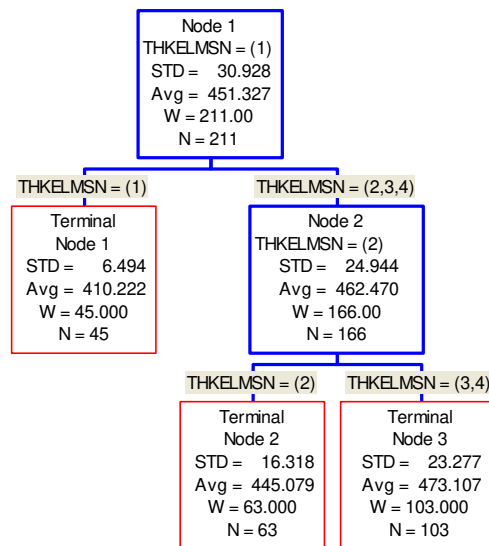
Keterangan: \*\* Optimal

Pada Tabel 4 terlihat bahwa dari 27 tujuh sub pohon yang terbentuk, sub pohon nomor 25 adalah pohon regresi optimal, karena nilai kesalahan relatif validasi silang dari sub pohon nomor 25 memiliki nilai paling kecil. Pohon regresi optimal yang terpilih tersebut seperti terlihat pada Gambar 2 memiliki 2 simpul dalam dan 3 simpul terminal, dengan kedalaman pohon regresi sebesar 3 tingkatan. Pada pohon regresi optimal yang terbentuk, tidak semua variabel prediktor yang diduga mempengaruhi jumlah indek mutu benang 30/1 Rayon masuk ke dalam model. Dari empat variabel prediktor hanya satu variabel prediktor yang masuk ke dalam model pohon regresi optimal yaitu variabel tahun keluaran mesin.

**Tabel 5.** Skor Variabel Penting pada Pohon Regresi Optimal

Variabel	Skor
Tahun keluaran mesin	100.00
Bahan Baku	4.55
Temperatur Ruangan	0.00
Kelembaban udara	0.00

Tabel 5 menunjukkan bahwa dalam pohon regresi optimal variabel tahun keluaran mesin juga memiliki nilai skor tertinggi seperti halnya pada pohon regresi maksimal, yang berarti variabel tahun keluaran mesin paling sering digunakan sebagai pemilah utama maupun pemilah pengganti pada pohon regresi optimal. Jadi, variabel tahun keluaran mesin merupakan variabel yang mempunyai kontribusi terbesar dalam pembentukan pohon regresi optimal. Variabel bahan baku, temperatur ruangan dan kelembaban udara (RH) memiliki skor yang cukup kecil karena jarang digunakan sebagai pemilah, sehingga tidak memberikan kontribusi yang berarti dalam pembentukan pohon regresi optimal.

**Gambar 2.** Pohon Regresi Optimal

### 3.5. Ketelitian Dugaan Pohon Regresi

Pohon regresi yang baik adalah pohon yang mampu memprediksi data atau amatan baru. Untuk tujuan tersebut maka sebanyak 11 data *testing* dimasukkan ke dalam pohon regresi optimal yang terbentuk untuk melakukan pengukuran ketelitian dugaan, yaitu melihat kemampuan pohon regresi yang terbentuk dalam menduga atau memprediksi besar jumlah indek mutu benang untuk data *testing*. Pada Tabel 6 dapat dilihat nilai prediksi dari amatan data *testing*. Sedangkan tingkat ketelitiannya adalah:

$$MSE = \frac{1}{11} \sum_{i=1}^{11} (Y_i - \hat{Y}_i)^2 = 86.347$$

Pohon Regresi optimal yang terbentuk seperti terlihat pada Gambar 2 didapat beberapa aturan klasifikasi yaitu:

1. Jika THKELMSN = (1) maka MEAN = 410.222.  
Berarti bahwa ketika benang 30/1 Rayon diproduksi oleh mesin Plat (Tahun Keluaran 1961) rata-rata nilai indek mutu benangnya adalah 410.222.
2. Jika THKELMSN = (2) maka MEAN = 445.079.

Berarti bahwa ketika benang 30/1 Rayon diproduksi oleh mesin Hawa Lama (Tahun Keluaran 1979) rata-rata nilai indek mutu benangnya adalah 445.079.

3. Jika THKELMSN = (3) dan THKELMSN = (4) maka MEAN = 473.107.

Berarti bahwa ketika benang 30/1 Rayon diproduksi oleh mesin Zinser dan Hawa Baru (Tahun Keluaran 1995 dan 1997) rata-rata nilai indek mutu benang yang diperoleh adalah 473.107.

**Tabel 6.** Perbandingan Data Pengamatan dan Data Prediksi

No	IMB	Prediksi
1	460	445.07937
2	410	410.22223
3	490	473.10678
4	480	473.10678
5	440	445.07938
6	410	410.22223
7	480	473.10678
8	470	473.10678
9	440	445.07938
10	410	410.22223
11	490	473.10678

#### 4. Kesimpulan

Berdasarkan dari analisis metode pohon regresi variabel yang berpengaruh terhadap jumlah indek mutu benang 30/1 Rayon di PT Industri Sandang Nusantara adalah variabel tahun keluaran mesin sebagai faktor utama dengan skor tertinggi sebesar 100. Sedangkan variabel bahan baku, kelembaban udara dan temperatur ruangan dengan skor variabel penting dibawah 20, sehingga secara statistik pengaruhnya dapat diabaikan. Pohon regresi optimal yang terbentuk terdiri dari 2 simpul dalam dan 3 simpul terminal, dengan kedalaman pohon regresi sebesar 2 tingkatan. Pemilahan pada simpul 1 dilakukan oleh variabel tahun keluaran mesin, dan pada simpul 2 juga dilakukan oleh variabel tahun keluaran mesin. Dari interpretasi pohon regresi optimal dapat dilihat bahwa pada variabel tahun keluaran mesin yang terbaru menghasilkan nilai indek mutu benang tertinggi.

#### DAFTAR PUSTAKA

- 1 Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., *Classification and Regression Trees*, Chapman & Hall, New York, 1993.
- 2 Komalasari, W.B., Metode Pohon Regresi untuk Ekploratori Data dengan Variabel yang Banyak dan Kompleks, *Jurnal Informatika Pertanian*, 2007, Vol. 16: 967-980.
- 3 Lewis, R.J., An Introduction to Classification and Regression Tree (CART) Analysis, *Presented at the 2000 Anual Meeting Of Society For Academy Emergency Medice in San Francisco, California*. 2000, URL: <http://www.saem.org/download/lewis1.pdf>
- 4 Steinberg, D. and Colla, P., *CART – Classification and Regression Trees*, CA, Salford System, San Diego, 1997.
- 5 Therneau, T.M. and Atkinson, E. J., An Introduction to Recursive Partitioning Using the RPART Routine. Thechnical Report 61, *Section of Biostatistics*, Mayo Clinic, Rochester, 1997, URL: <http://www.mayo.edu/har/techrpt/61.pdf>